

АКАДЕМИЯ НАУК СССР
НАУЧНЫЙ ЦЕНТР БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР

ПРЕПРИНТ

В.Ю.ЛУНИН

**ОПТИМИЗАЦИЯ РАСЧЕТА
СТРУКТУРНЫХ ФАКТОРОВ
В КРИСТАЛЛОГРАФИИ БЕЛКА**

ПУЩИНО • 1982

УДК 548.73

Рассматривается вопрос о сокращении затрат машинного времени на расчет набора структурных факторов при кристаллографическом уточнении белковой структуры. Получены приближенные оценки, позволяющие выбирать параметры метода так, чтобы обеспечить минимальное время работы при сохранении заданной точности.

Работа будет направлена для опубликования в журнал "Кристаллография".

The problem of reducing of computer requirements for calculation of structure factors for large molecules was considered. The approximations were obtained for optimal parameters.

ВВЕДЕНИЕ

1. При кристаллографическом уточнении структуры белка в обратном пространстве (см., например, /1/) одним из этапов работы является расчет набора структурных факторов (коэффициентов Фурье) функции распределения электронной плотности по некоторой модели структуры. При работе с белками при высоком и среднем разрешении этот этап является довольно трудоемким и во многом определяет скорость работы программы уточнения структуры. Модули рассчитанных структурных факторов, как правило, используются для вычисления некоторого критерия близости модельной и истинной структур, который обычно конструируется из разностей вычисленных по модели и измеренных экспериментально модулей структурных факторов. Из такого вида критерия следует, что не имеет смысла рассчитывать структурные факторы по модели с точностью, значительно превышающей точность экспериментального измерения модулей структурных факторов (например, в /6/ эта точность менялась от 4% при разрешении 3,5 Å до 14% при разрешении 2,8 Å).

Таким образом, возникает задача, как построить алгоритм вычисления структурных факторов, чтобы при обеспечении заданной точности расчета минимизировать время вычислений. Рассмотрению одного из путей практического решения этой задачи и посвящена данная работа.

Автор признателен А.Г.Уржумцеву и Е.А.Вернословой за обсуждения работы и помощь в проведении тестов.

2. Далее в работе будут использоваться следующие обозначения:

$\bar{a}, \bar{b}, \bar{c}$ – периоды кристаллической решетки;

$\bar{a}^*, \bar{b}^*, \bar{c}^*$ – периоды решетки в "обратном" пространстве;

B_j^t – температурный фактор (изотропный) j -го атома;

B^c – "добавочное значение" температурного фактора;

$B_{q,j}, C_{q,j}$ – параметры, определяющие фактор атомного рассеяния;

$\vec{F}(\vec{h}) = F(\vec{h}) e^{i\varphi(\vec{h})}$ - структурный фактор, соответствующий точке \vec{h} обратного пространства;

$q_j(\vec{h})$ - фактор атомного рассеяния j -го атома;

M - число атомов в элементарной ячейке;

N - общее число рассчитываемых структурных факторов;

n_x, n_y, n_z - число делений по осям элементарной ячейки при выборе сетки, на которой рассчитывается функция распределения электронной плотности;

$\vec{R}_j = (x, y, z) = x\vec{a} + y\vec{b} + z\vec{c}$ - вектор из начала координат в центр j -го атома;

$R = |\vec{R}|$ - длина вектора \vec{R} ;

\mathcal{R} - решетка в "прямом" пространстве - совокупность точек вида $u\vec{a} + v\vec{b} + w\vec{c}$; u, v, w - целые числа;

\mathcal{R}' - решетка в "обратном" пространстве - совокупность точек вида $h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$; h, k, l - целые числа;

$\vec{r} = (h, k, l) = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$ - вектор из начала координат в точку решетки \mathcal{R}' ;

$r = |\vec{r}|$ - длина вектора \vec{r} ;

V - элементарная ячейка - параллелепипед, построенный на векторах $\vec{a}, \vec{b}, \vec{c}$;

$|V|$ - объем элементарной ячейки;

Z - множество целых чисел;

$\rho_j(\vec{R})$ - функция распределения электронной плотности в j -м атоме структуры;

$\rho_{0j}(R)$ - сферически симметричная функция распределения электронной плотности в j -м атоме структуры, сдвинутом в начало координат;

$\rho_{0j}^{\times}(R)$ - функция $\rho_{0j}(R)$ замененная нулем при $R > \times$;

$\rho(\vec{R})$ - функция распределения электронной плотности в кристалле;

$\tilde{\rho}(\vec{R})$ - модифицированная функция $\rho(\vec{R})$ (см. § 2, п. 6);

τ - шаг сетки, на которой рассчитывается функция $\tilde{\rho}(\vec{R})$.

§ 1. Алгоритм расчета структурных факторов

1. Пусть $\{\vec{a}, \vec{b}, \vec{c}\}$ - базис, построенный на периодах кристаллической решетки и $\{\vec{a}^*, \vec{b}^*, \vec{c}^*\}$ - сопряженный базис. Обозначим $\mathcal{R} = \{m\vec{a} + n\vec{b} + l\vec{c}\}_{m,n,l \in Z}$ - целочисленную решетку, построенную на векторах $\vec{a}, \vec{b}, \vec{c}$. Пусть в элементарной ячейке V содержится M атомов, каждый из которых предполагает-

ся сферически симметричным и характеризуется типом атома, координатами центра $\vec{R}_j = (x_j, y_j, z_j) = x_j \vec{a} + y_j \vec{b} + z_j \vec{c}$ и индивидуальным температурным фактором B_j^t ($j=1, \dots, M$). Тогда для точки "обратной" решетки $\vec{k} = (h, k, l) = h \vec{a}^* + k \vec{b}^* + l \vec{c}^*$ (h, k, l - целые числа) соответствующий структурный фактор $\vec{F}(\vec{k})$ дается формулой (см. /1/):

$$\vec{F}(\vec{k}) = F(\vec{k}) e^{i\varphi(\vec{k})} = \sum_{j=1}^M q_j(\vec{k}) e^{-t B_j^t k^2} e^{2\pi i (\vec{k}, \vec{R}_j)} \quad (1)$$

где $q_j(\vec{k})$ - фактор атомного рассеяния, который зависит лишь от сорта атома и предполагается для каждого атома известным.

При этом обычно требуется рассчитать все структурные факторы $\vec{F}(\vec{k})$ с $k_{\min} \leq k \leq k_{\max}$, где k_{\min}, k_{\max} - заданные значения.

Идейно наиболее простой путь вычисления величин $\vec{F}(\vec{k})$ - это непосредственный расчет по формуле (1). Однако за эту простоту приходится расплачиваться большим объемом вычислений: структура формулы (1) такова, что объем вычислений пропорционален MN , где N - общее число рассчитываемых структурных факторов, и при тысячах атомов и десятках тысяч рассчитываемых структурных факторов время расчета становится значительным.

Использование свойств симметрии структуры и различные "вычислительные хитрости" позволяют преобразовать формулу (1) к менее трудоемкому виду. Однако, как показывает анализ, проведенный Тен Эйком в работе /3/, более существенного выигрыша во времени удается достичь, применив "обходной" путь расчета величин $\vec{F}(\vec{k})$.

2. Пусть распределение электронной плотности j -го атома задается функцией $\rho_j(\vec{r})$, тогда функция распределения электронной плотности в кристалле

$$\rho(\vec{R}) = \sum_{\vec{u} \in \mathcal{R}} \sum_{j=1}^M \rho_j(\vec{R} - \vec{u}) \quad (2)$$

и искомые структурные факторы (см. /1/):

$$\vec{F}(\vec{k}) = \int_V \rho(\vec{r}) e^{2\pi i (\vec{k}, \vec{r})} dV_{\vec{r}} = |V| \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) e^{2\pi i (hx + ky + lz)} dx dy dz \quad (3)$$

где V - параллелепипед, построенный на векторах $\vec{a}, \vec{b}, \vec{c}$, и $|V|$ - его объем.

Обсуждаемый способ расчета величин $\vec{F}(\vec{k})$ состоит из двух

этапов: а) расчет на некоторой сетке в элементарной ячейке функции $\rho(x, y, z)$ согласно формуле (2); б) выполнение преобразования (3).

Функции $\rho_j(\vec{R})$ быстро убывают при удалении от центра атома, поэтому при вычислении функции $\rho(x, y, z)$ в некоторой точке мы можем ограничиться лишь суммированием вкладов от нескольких близлежащих атомов. Следовательно, объем вычислений на этом этапе пропорционален числу атомов и не зависит от числа узлов сетки.

Применяя на этапе б) для вычисления интеграла простейшую квадратурную формулу:

$$\vec{F}(\vec{\lambda}) \cong \frac{|V|}{n_x n_y n_z} \sum_{u=0}^{n_x-1} \sum_{v=0}^{n_y-1} \sum_{w=0}^{n_z-1} \rho\left(\frac{u}{n_x}, \frac{v}{n_y}, \frac{w}{n_z}\right) e^{2\pi i (h \frac{u}{n_x} + k \frac{v}{n_y} + l \frac{w}{n_z})} \quad (4)$$

мы можем использовать при расчете алгоритм быстрого преобразования Фурье, который требует порядка $N \log N$ операций.

Использование свойств симметрии структуры может быть использовано на этом этапе для сокращения времени вычислений (см. /4/).

Таким образом, суммарная "цена вычислений" при расчете по обсуждаемому алгоритму $C_1 M + C_2 N \log N$ при большом числе атомов оказывается ниже "цены" $C_3 MN$ непосредственного расчета по формуле (1).

3. Для функции распределения электронной плотности j -го атома мы имеем $\rho_j(\vec{R}) = \rho_{0j}(\vec{R} - \vec{R}_j)$, где $\rho_{0j}(\vec{R})$ - распределение электронной плотности для атома, находящегося в начале координат. При этом для сферически симметричной функции

$\rho_{0j}(\vec{R}) = \rho_{0j}(R)$ связь с функцией $q_j(\lambda)$, входящей в выражение (1), дается равенствами

$$q_j(\lambda) e^{-\frac{1}{4} B_j^{\dagger} \lambda^2} = \int_V \sum_{\vec{u} \in \mathcal{R}} \rho_{0j}(|\vec{R} - \vec{u}|) e^{2\pi i (\vec{\lambda}, \vec{R})} dV_{\vec{R}} =$$

$$= \int_{\mathbb{R}^3} \rho_{0j}(R) e^{2\pi i (\vec{\lambda}, \vec{R})} dV_{\vec{R}} = \frac{2}{\lambda} \int_0^{\infty} R \rho_{0j}(R) \sin(2\pi \lambda R) dR$$

и

$$\rho_{0j}(R) = \frac{2}{R} \int_0^{\infty} \lambda q_j(\lambda) e^{-\frac{1}{4} B_j^{\dagger} \lambda^2} \sin(2\pi \lambda R) d\lambda.$$

Функцию $q_j(\lambda)$ обычно аппроксимируют суммой гауссовых функций (см., например, /5/):

$$q_j(\lambda) = \sum_{q=1}^P C_q e^{-\frac{1}{4} B_{qj} \lambda^2}, \quad (5)$$

где p , как правило, равно 1 или 2 (реже - 3). Поскольку

$$\frac{2}{R} \int_0^{\infty} e^{-\frac{1}{2}B\lambda^2} \lambda \sin(2\pi\lambda R) d\lambda = \left(\frac{4\pi}{B}\right)^{3/2} e^{-4\pi^2 R^2/B}$$

то тем самым распределение электронной плотности атома также аппроксимируется суммой гауссовых функций:

$$\rho_{0j}(R) = \sum_{q=1}^p C_{q,j} \left(\frac{4\pi}{B_{q,j} + B_j^t}\right)^{3/2} e^{-\frac{4\pi^2 R^2}{B_{q,j} + B_j^t}} \quad (6)$$

Функции $\rho_{0j}(R)$ быстро убывают с ростом R , поэтому при расчете функции распределения электронной плотности в кристалле

$\rho(\vec{R})$, мы будем полагать, что распределение электронной плотности отдельного атома описывается функцией $\rho_{0j}^{\alpha_j}(|\vec{R} - \vec{R}_j|)$, где

$$\rho_{0j}^{\alpha_j}(R) = \begin{cases} \rho_{0j}(R) & \text{при } R \leq \alpha_j, \\ 0 & \text{при } R > \alpha_j. \end{cases} \quad (7)$$

При этом мы будем допускать разный "радиус" α_j для разных атомов.

Таким образом, выражение (2) для функции $\rho(\vec{R})$ приобретает вид:

$$\rho(\vec{R}) = \sum_{j=1}^M \sum_{\vec{u} \in Q} \rho_{0j}^{\alpha_j}(|\vec{R} - \vec{u} - \vec{R}_j|),$$

где в правой части при любом \vec{R} стоит лишь конечное число ненулевых слагаемых.

4. Ясно, что сокращение времени вычислений путем выполнения преобразования (4) на грубой сетке приводит к возрастанию погрешности в вычислении величин $\vec{F}(\vec{\lambda})$ (по сравнению с их теоретическими значениями, дающимися формулой (1)). Для уменьшения этой погрешности при неизменной стеке может быть применено введение "искусственного температурного фактора" (см. /3/).

Суть этого приема состоит в следующем. Ошибка, совершаемая при замене интеграла (2) по формуле (4) будет, вообще говоря, тем больше, чем более резкие пики имеет подынтегральная функция (математически это выражается в оценке погрешности формулы (4) через вторые производные подынтегральной функции). Для описывающих атом функций типа $\exp(-4\pi^2 R^2/B)$ соответствующий пик будет тем резче, чем меньше значение B . Поэтому ошибки, вносимые в структурные факторы на этом этапе, будут тем меньше, чем более "расплывшимися" будут функции, описывающие плотность отдельных атомов, то есть чем выше будут значения температурных факторов отдельных атомов.

Свойство преобразования Фурье переводить свертку функций в произведение Фурье-образов позволяет модифицировать равенство (3) следующим образом. Пусть $\varphi(\vec{R})$ - некоторая функция, тогда

$$\begin{aligned}\vec{F}(\vec{\lambda}) \cdot \vec{\Phi}(\vec{\lambda}) &= \int_{R^3} \left[\sum_{j=1}^M (\rho_{0j} * \varphi)(\vec{R} - \vec{R}_j) \right] \exp(2\pi i(\vec{\lambda}, \vec{R})) dV_{\vec{R}} = \\ &= \int_{V} \left[\sum_{j=1}^M \sum_{\vec{u} \in \mathcal{R}} (\rho_{0j} * \varphi)(\vec{R} - \vec{R}_j - \vec{u}) \right] \exp(2\pi i(\vec{\lambda}, \vec{R})) dV_{\vec{R}},\end{aligned}$$

где $\vec{\Phi}(\vec{\lambda}) = \int_{R^3} \varphi(\vec{R}) \exp(2\pi i(\vec{\lambda}, \vec{R})) dV_{\vec{R}}$,

$$(\rho_{0j} * \varphi)(\vec{R}) = \int_{R^3} \rho(\vec{u}) \varphi(\vec{R} - \vec{u}) dV_{\vec{u}}.$$

Если теперь взять $\varphi(\vec{R}) = (4\pi/B^3)^{3/2} \exp(-4\pi^2 R^2/B^3)$, то $\vec{\Phi}(\vec{\lambda}) = \exp(-\frac{1}{4} B^3 \lambda^2)$, и если $\rho_{0j}(R)$ задается в виде (6), то

$$(\rho_{0j} * \varphi)(\vec{R}) = \sum_{q=1}^P \left(\frac{4\pi}{B_{q,j} + B_j^t + B^a} \right)^{3/2} \exp\left(-\frac{4\pi^2 R^2}{B_{q,j} + B_j^t + B^a}\right), \quad (8)$$

то есть подынтегральная функция

$$\tilde{\rho}_{B^a}(R) = \sum_{j=1}^M \sum_{\vec{u} \in \mathcal{R}} (\rho_{0j} * \varphi)(\vec{R} - \vec{R}_j - \vec{u})$$

снова является суммой гауссовых функций и при всех B^a

$$\vec{F}(\vec{\lambda}) = \exp\left(\frac{1}{4} B^a \lambda^2\right) \int_V \tilde{\rho}_{B^a}(\vec{R}) \exp(2\pi i(\vec{\lambda}, \vec{R})) dV_{\vec{R}}. \quad (9)$$

В силу (8) преобразование формулы (3) к виду (9) формально эквивалентно тому, что мы рассчитываем структурные факторы для структуры, все атомы которой имеют увеличенное на одну и ту же величину B^a значения температурных факторов, а затем умножаем эти структурные факторы на множители $\exp(\frac{1}{4} B^a \lambda^2)$.

Поскольку за счет добавочного значения B^a форма отдельных атомов "расплывается", то это приводит к уменьшению относительной ошибки, вносимой заменой интеграла на конечную сумму.

Отметим, что внесение добавочного вклада B^a в температурный фактор практически не влияет на время расчета набора структурных факторов.

5. Не следует, однако, думать, что мы можем безнаказанно сколь угодно увеличивать значения B^a . Источниками ошибки при расчете структурных факторов по излагаемой методике служат как

замена интеграла (3) на сумму (4), так и обрыв электронной плотности отдельного атома на расстоянии \mathcal{R}_j от центра в формуле (7). Если ошибки, связанные с использованием формулы (4), убывают при введении добавочного значения B^a в температурный фактор (то есть при "расплывании" формы атома), то ошибки, вызванные обрывом функции $\rho_{0j}(R)$ при $R = \mathcal{R}_j$, наоборот, при этом возрастают. Поэтому возникает вопрос о выборе оптимального значения B_{opt}^a так, чтобы итоговая точность расчета структурных факторов (определяемая как подробность сетки, на которой ведется расчет, так и "радиусом" атомов) была при заданных параметрах сетки и радиусах атомов максимальна. Этот вопрос будет более подробно рассмотрен в следующих параграфах.

6. Итак, далее везде будет подразумеваться, что расчет структурных факторов $\vec{F}(\vec{h})$ с $h_{min} \leq h \leq h_{max}$ ведется по следующему алгоритму:

а) на сетке $\left\{ \frac{u}{n_x}, \frac{v}{n_y}, \frac{w}{n_z} \right\}$, $u=0, \dots, n_x-1$; $v=0, \dots, n_y-1$; $w=0, \dots, n_z-1$ рассчитывается функция

$$\tilde{\rho}(R) = \sum_{j=1}^M \sum_{\vec{u} \in R} \rho_{0j}^{\mathcal{R}_j}(|R - \vec{R}_j - \vec{u}|), \quad (10)$$

где

$$\rho_{0j}^{\mathcal{R}_j}(R) = \begin{cases} \sum_{q=1}^P C_{q,j} \left(\frac{4\pi}{B_{q,j} + B_j^t + B^a} \right)^{3/2} \exp\left(\frac{-4\pi^2 R^2}{B_{q,j} + B_j^t + B^a} \right) & \text{при } R < \mathcal{R}_j, \\ 0 & \text{при } R > \mathcal{R}_j \end{cases}$$

и фактически суммирование в (10) ведется по всем атомам кристалла, центры которых либо находятся внутри элементарной ячейки V , либо удалены от нее на расстояние не выше \mathcal{R}_j (то есть для некоторых атомов вклад в сумму (10) вносят и трансляционно связанные с ними атомы);

б) выполняется преобразование

$$\vec{F}(h, k, l) = \exp\left(\frac{1}{4} B^a h^2 \right) \frac{V}{n_x n_y n_z} \sum_{u=0}^{n_x-1} \sum_{v=0}^{n_y-1} \sum_{w=0}^{n_z-1} \tilde{\rho}\left(\frac{u}{n_x}, \frac{v}{n_y}, \frac{w}{n_z} \right) \exp\left\{ 2\pi i \left(h \frac{u}{n_x} + k \frac{v}{n_y} + l \frac{w}{n_z} \right) \right\}.$$

Отметим, что при этом параметры $\vec{R}_j, B_{q,j}, C_{q,j}, B_j^t$

для каждого атома предполагаются заданными, в то время как выбор параметров $n_x, n_y, n_z, \mathcal{R}_j$ и B^a лежит на нас. В

следующих параграфах мы обсудим вопрос о том, как выбирать эти параметры, чтобы обеспечить желаемую точность расчета величин $\vec{F}(\vec{h})$ за минимальное время.

§ 2. Влияние "радиуса" атома на точность
расчета структурных факторов

1. В этом параграфе будет предложен способ грубой оценки погрешности, вносимой в рассчитанные структурные факторы "обрывом" функции распределения электронной плотности в атоме. То есть мы попытаемся ответить на вопрос: как скажется на точности расчета величин $\vec{F}(\vec{\lambda})$ использование в равенстве (10) функций ρ_{0j}^{∞} вместо ρ_{0j} ?

В реальной структуре мы имеем дело с разными атомами и разными значениями температурного фактора B_j^t . Однако основные атомы (C, N, O), составляющие белковую молекулу, имеют сходные факторы атомного рассеяния, и мы примем потому упрощающее предположение, что все атомы в структуре одинаковы и фактор атомного рассеяния (5) аппроксимируется суммой двух гауссовых функций. В качестве параметров аппроксимации мы возьмем параметры атома азота согласно /5/:

$$\begin{aligned} B_{1,j} &= B_1 = 25.0, & B_{2,j} &= B_2 = 3.41; \\ C_{1,j} &= C_1 = 3.05, & C_{2,j} &= C_2 = 3.94. \end{aligned} \quad (11)$$

Кроме того, мы будем считать, что все атомы имеют одинаковое значение температурного фактора B^t (равное среднему значению температурного фактора для атомов структуры).

Предполагая, что в расчет введен "добавочный температурный фактор" B^a , мы можем записать функцию $\rho_{0j}(R)$ в виде:

$$\rho_{0j}(R) = \rho_0(R) = C_1 \left(\frac{4\pi}{B_1+B} \right)^{3/2} \exp\left(-\frac{4\pi^2 R^2}{B_1+B}\right) + C_2 \left(\frac{4\pi}{B_2+B} \right)^{3/2} \exp\left(-\frac{4\pi^2 R^2}{B_2+B}\right),$$

где $B = B^t + B^a$.

2. При отсутствии "обрыва" функции $\rho_0(R)$ структурные факторы такой структуры имели бы вид:

$$\vec{F}(\vec{\lambda}) = \left[\frac{2}{\lambda} \int_0^{\infty} R \rho_0(R) \sin(2\pi \lambda R) dR \right] \sum_{j=1}^M e^{2\pi i(\vec{\lambda}, \vec{R}_j)}.$$

Замена же функции $\rho_0(R)$ нулем при $R \geq \infty$ приведет к тому, что мы получим вместо $\vec{F}(\vec{\lambda})$ величины

$$\vec{F}_{\infty}(\vec{\lambda}) = \left[\frac{2}{\lambda} \int_0^{\infty} R \rho_0(R) \sin(2\pi \lambda R) dR \right] \sum_{j=1}^M e^{2\pi i(\vec{\lambda}, \vec{R}_j)},$$

что приводит к ошибке

$$\frac{|\vec{F}(\vec{\lambda}) - \vec{F}_{\infty}(\vec{\lambda})|}{|\vec{F}(\vec{\lambda})|} = 2 \left\{ C_1 \exp\left[-\frac{1}{4}(B_1+B)\lambda^2\right] + C_2 \exp\left[-\frac{1}{4}(B_2+B)\lambda^2\right] \right\}^{-1} \times$$

$$\times \int_{\alpha}^{\infty} R \left[C_1 \left(\frac{4\pi}{B_1+B} \right)^{3/2} \exp\left(-\frac{4\pi^2 R^2}{B_1+B}\right) + C_2 \left(\frac{4\pi}{B_2+B} \right)^{3/2} \exp\left(-\frac{4\pi^2 R^2}{B_2+B}\right) \right] \sin(2\pi R) dR.$$

Мы упростим эту оценку, заменив в подынтегральном выражении $\sin(2\pi R)$ на единицу и оставив в квадратных скобках по одной (более медленно убывающей) экспоненте. В таком случае, выполнив интегрирование, мы получим для отдельного структурного фактора оценку погрешности в виде:

$$\frac{|\vec{F}(\vec{h}) - \vec{F}_{\alpha}(\vec{h})|}{|\vec{F}(\vec{h})|} \approx \frac{1}{h C_2 \sqrt{\pi(B_1+B)}} \exp\left[\frac{1}{4}(B_2+B)h^2\right] \exp\left(-\frac{4\pi^2 \alpha^2}{B_1+B}\right). \quad (12)$$

Предполагая, что нам надо рассчитать все структурные факторы $\vec{F}(\vec{h})$ с $|\vec{h}| \leq h_{\max}$, мы введем в рассмотрение функцию

$U(\alpha, B, h_{\max})$, характеризующую среднюю погрешность набора рассчитанных структурных факторов, вызванную конечным радиусом атома, при радиусе атома α и заданном значении $B = B_1 + B_2$:

$$U(\alpha, B, h_{\max}) = \frac{3}{4\pi h_{\max}^3} \int_{|\vec{h}| \leq h_{\max}} \frac{|\vec{F}(\vec{h}) - \vec{F}_{\alpha}(\vec{h})|}{|\vec{F}(\vec{h})|} dV_{\vec{h}}.$$

Применяя полученную аппроксимацию (12) подынтегрального выражения и проводя интегрирование, получим

$$U(\alpha, B, h_{\max}) \approx \frac{12C_1}{C_2} \frac{1}{(B_2+B)\sqrt{\pi(B_1+B)}} \cdot \frac{1}{h_{\max}^3} \left\{ \exp\left[\frac{1}{4}(B_2+B)h_{\max}^2\right] - 1 \right\} \exp\left(-\frac{4\pi^2 \alpha^2}{B_1+B}\right) \quad (13)$$

3. Для проверки точности полученной оценки был проведен контрольный расчет структурных факторов для структуры актинидина (координаты были взяты из банка белковых молекул) непосредственно по формуле (1). При этом величины n_x, n_y, n_z брались достаточно большими, чтобы ошибка, вызванная заменой интеграла (3) на сумму (4), была несущественна. Результаты сравнения реально достигнутой точности с оценкой точности (13) приведены в табл. 1.

Т а б л и ц а 1

Соответствие реальной и оценочной точности

α	(Å)	Реальная точность*	Оценка точности**	
2,15		0,03	0,02	$\lambda_{\max} = 0,5;$
1,71		0,10	0,07	$B = 20$
2,47		0,011	0,004	

* Величина $\langle |\hat{F} - \bar{F}_\alpha| / |\hat{F}| \rangle$ по 50 случайным образом выбранным рефлексам.

** Значение величины $U(\alpha, B, \lambda_{\max})$ согласно формуле (13).

§ 3. Влияние величины шага сетки на точность расчета структурных факторов

1. В этом параграфе мы дадим оценку погрешности, вносимой в рассчитанные структурные факторы заменой интеграла

$$\vec{F}(\vec{h}) = |\nu| \iiint \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dx dy dz$$

на сумму

$$\vec{F}_\tau(\vec{h}) = \frac{|\nu|}{n_x n_y n_z} \sum_{u=0}^{n_x-1} \sum_{v=0}^{n_y-1} \sum_{w=0}^{n_z-1} \rho\left(\frac{u}{n_x}, \frac{v}{n_y}, \frac{w}{n_z}\right) e^{2\pi i(h\frac{u}{n_x} + k\frac{v}{n_y} + l\frac{w}{n_z})}.$$

При этом мы предположим, что числа делений n_x, n_y, n_z выбраны так, что шаг сетки одинаков по всем направлениям:

$$\tau = \frac{a}{n_x} \approx \frac{b}{n_y} \approx \frac{c}{n_z}. \quad (14)$$

Можно показать (см., например, /3/), что

$$\vec{F}_\tau(\vec{h}) = \vec{F}(\vec{h}) + \sum_{\substack{u, v, w \in \mathbb{Z} \\ (u, v, w) \neq 0}} \vec{F}(\vec{h} + u n_x \vec{a}^* + v n_y \vec{b}^* + w n_z \vec{c}^*), \quad (15)$$

то есть абсолютная величина ошибки определяется суммой "точных" структурных факторов $\vec{F}(\vec{h} + \vec{h}_j)$ по всем ненулевым векторам \vec{h}_j решетки $\{u(n_x \vec{a}^*) + v(n_y \vec{b}^*) + w(n_z \vec{c}^*)\}_{u, v, w \in \mathbb{Z}}$.

Мы будем по-прежнему считать, что все атомы в структуре одинаковы и моделируются атомом азота, так что

$$\vec{F}(\vec{h}) = f(N) \sum_{j=1}^M e^{2\pi i(\vec{h}, \vec{R}_j)},$$

где

$$f(\lambda) = C_1 \exp[-\frac{1}{4}(B_1+B)\lambda^2] + C_2 \exp[-\frac{1}{4}(B_2+B)\lambda^2],$$

$B = B^t + B^a$, и константы C_1, C_2, B_1, B_2 определены в (11).

Ясно, что в сумму (15) существенный вклад вносит лишь небольшое число слагаемых, отвечающих ближайшим к началу координат точкам $\vec{\lambda}_j$. Мы предположим для простоты, что оси ячейки ортогональны, тогда в силу условия (14) ближайшие к началу координат точки - это:

$$\vec{\lambda}_1 = n_x \vec{a}^*, \vec{\lambda}_2 = -n_x \vec{a}^*, \vec{\lambda}_3 = n_y \vec{b}^*, \vec{\lambda}_4 = -n_y \vec{b}^*, \vec{\lambda}_5 = n_z \vec{c}^*, \vec{\lambda}_6 = -n_z \vec{c}^*.$$

При этом все они находятся на расстоянии $|\vec{\lambda}_j| = \frac{1}{l}$ от начала координат.

Таким образом, мы оценим относительную погрешность величиной:

$$\frac{|\vec{F}(\vec{\lambda}) - \vec{F}_r(\vec{\lambda})|}{|\vec{F}(\vec{\lambda})|} \approx \frac{|\sum_{j=1}^6 \vec{F}(\vec{\lambda} + \vec{\lambda}_j)|}{|\vec{F}(\vec{\lambda})|} \quad (16)$$

2. Следующий шаг в выводе оценки будет заключаться в замене правой части равенства (16) (зависящей от координат атомов структуры) на ее среднее по всевозможным структурам значение.

Мы рассмотрим координаты атомов (x_j, y_j, z_j) , $j=1, \dots, M$ как независимые случайные величины, равномерно распределенные на отрезке $[0, 1]$. В таком случае (см. /2/) случайные величины $\eta = |\vec{F}(\vec{\lambda})|$ и $\xi = |\sum_{j=1}^6 \vec{F}(\vec{\lambda} + \vec{\lambda}_j)|$ имеют плотности распределения вероятностей

$$P_\eta(\eta) = \frac{2\eta}{\sigma_\eta^2} \exp(-\frac{\eta^2}{\sigma_\eta^2}), \quad P_\xi(\xi) = \frac{2\xi}{\sigma_\xi^2} \exp(-\frac{\xi^2}{\sigma_\xi^2}),$$

где

$$\sigma_\eta^2 = M f^2(\lambda), \quad \sigma_\xi^2 = M \sum_{j=1}^6 f^2(\vec{\lambda} + \vec{\lambda}_j)$$

Рассматривая ξ и η как независимые случайные величины, мы будем иметь для случайных величин $u = \xi \eta$ и $v = \xi / \eta$ (см.

/6/) совместное распределение вероятностей

$$P(u, v) = \frac{2u}{\sigma_\xi^2 \sigma_\eta^2 v} \exp(-\frac{uv}{\sigma_\xi^2} - \frac{v}{\sigma_\eta^2} u).$$

Поэтому

$$\langle \frac{\xi}{\eta} \rangle = \int_0^\infty \int_0^\infty v P(u, v) dv du = \frac{2}{\sigma_\xi^2 \sigma_\eta^2} \int_0^\infty \int_0^\infty u \exp[-(\frac{v}{\sigma_\xi^2} + \frac{1}{\sigma_\eta^2} v)u] du dv =$$

$$= 2\sigma_{\vec{r}}^2 \sigma_{\vec{v}}^2 \int_0^{\infty} \frac{v^2}{(\sigma_{\vec{r}}^2 v^2 + \sigma_{\vec{r}}^2)^2} dv = \frac{\pi}{2} \cdot \frac{\sigma_{\vec{r}}^2}{\sigma_{\vec{v}}^2}.$$

Таким образом,

$$\left\langle \frac{\left| \sum_{j=1}^6 \vec{F}(\vec{h} + \vec{h}_j) \right|}{|\vec{F}(\vec{h})|} \right\rangle_{\{v_j\}} = \frac{\pi}{2} \frac{\sqrt{\sum_{j=1}^6 f^2(\vec{h} + \vec{h}_j)}}{f(\vec{h})},$$

и мы заменим оценку (16) на

$$\frac{|\vec{F}(\vec{h}) - \vec{F}_{\tau}(\vec{h})|}{|\vec{F}(\vec{h})|} \simeq \frac{\pi}{2} \frac{\sqrt{\sum_{j=1}^6 f^2(\vec{h} + \vec{h}_j)}}{f(\vec{h})}. \quad (17)$$

3. Предполагая, что нам надо рассчитать все структурные факторы $\vec{F}(\vec{h})$ с $|\vec{h}| \leq h_{\max}$, мы введем в рассмотрение функцию $V(\tau, B, h_{\max})$, характеризующую среднюю относительную погрешность набора рассчитанных структурных факторов

$$V(\tau, B, h_{\max}) = \frac{3}{4\pi h_{\max}^3} \int_{|\vec{h}| \leq h_{\max}} \frac{|\vec{F}(\vec{h}) - \vec{F}_{\tau}(\vec{h})|}{|\vec{F}(\vec{h})|} dV_{\vec{h}}.$$

Используя оценку (17) и оставляя в $f(\vec{h})$ только более "медленную" экспоненту, мы получим:

$$V(\tau, B, h_{\max}) \simeq \frac{\pi}{2} \exp\left(-\frac{B+B_2}{4\tau^2}\right) \cdot \frac{3}{4\pi h_{\max}^3} \int_{|\vec{h}| \leq h_{\max}} \sqrt{\sum_{j=1}^6 \exp[-(B+B_2)(\vec{h}, \vec{h}_j)]} dV_{\vec{h}}.$$

При ортогональных осях ячейки интеграл преобразуется к виду

$$V(\tau, B, h_{\max}) \simeq \frac{\pi}{2} \exp\left(-\frac{B+B_2}{4\tau^2}\right) \times \\ \times \frac{3\sqrt{2}}{4\pi h_{\max}^3} \int_{\sqrt{x^2+y^2+z^2} \leq h_{\max}} \sqrt{\text{ch}\left(\frac{B+B_2}{\tau}x\right) + \text{ch}\left(\frac{B+B_2}{\tau}y\right) + \text{ch}\left(\frac{B+B_2}{\tau}z\right)} dx dy dz.$$

Таким образом, приходим к оценке:

$$V(\tau, B, h_{\max}) \simeq \frac{\pi}{2} \exp\left(-\frac{B+B_2}{4\tau^2}\right) \Psi\left(\frac{B+B_2}{\tau} h_{\max}\right), \quad (18)$$

где

$$\Psi(t) = \frac{3\sqrt{2}}{4\pi t^3} \int_{\sqrt{x^2+y^2+z^2} \leq t} \sqrt{\text{ch}x + \text{ch}y + \text{ch}z} dx dy dz.$$

На рисунке приведен график функции $\psi = \ln \Psi(t)$. Отметим, что при реально встречающихся значениях аргумента функция $\Psi(t)$ с достаточной для нас точностью приближается, например, функцией

$$\psi = 0.77 \left(1 + \frac{1}{0.46 + 0.15 t^2}\right) e^{0.38 t}.$$

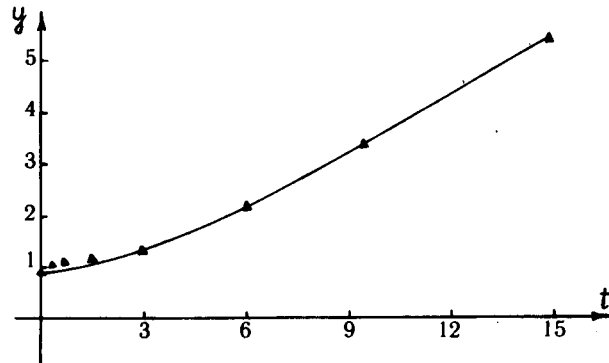


Рис. График функции $y = \ln \Psi(t)$ и значения
 $y = \ln \left\{ 0,77 \left(1 + \frac{t}{0,46 + 0,15t^2} \right) e^{0,38t} \right\}$

4. Для проверки точности полученной оценки был проведен контрольный расчет структурных факторов для структуры актинидина непосредственно по формуле (1). При этом радиус атома \mathcal{R} брался достаточно большим, чтобы ошибки, вызванные величиной \mathcal{R} , были несущественны. Результаты сравнения реально достигнутой точности с оценкой (18) приведены в табл. 2.

Т а б л и ц а 2

Соответствие реальной и оценочной (20) точности

τ (Å)	B	Реальная точность *	Оценка точности **	
0,6	0,5	0,30	0,39	
0,6	3,5	0,07	0,10	
0,6	9,5	0,01	0,02	
0,46	0,5	0,09	0,08	$\lambda_{max} = 0,5$
0,46	3,5	0,01	0,01	

* $\langle |\vec{F} - \vec{F}_r| / |\vec{F}| \rangle$ по 50 рефлексам.

** $V(\tau, B, \lambda_{max})$ согласно (18).

§ 4. Оптимальный выбор параметров метода

1. Из предыдущих рассмотрений следует, что при радиусе атома \mathcal{R} и шаге сетки τ среднюю относительную погрешность в структурных факторах можно охарактеризовать величиной

Т а б л и ц а 3

Значения наилучшей достижимой точности ϵ_{opt} и значения ν_{opt}
(в скобках), при которых она достигается для случая $\lambda_{max} = 0,5$

ϵ (Å) / ν (Å)	1,25	1,5	1,75	2,0	2,25	2,5	2,75	3,0
0,4	0,07 (2.)	0,03 (3.)	0,01 (4.)	0,003 (5.)	0,001 (6.)	-	-	-
0,5	0,09 (4.)	0,04 (5.)	0,02 (6.)	0,01 (8.)	0,003 (9.)	0,001 (10.)	-	-
0,6	0,12 (8.)	0,06 (9.)	0,03 (10.)	0,01 (12.)	0,01 (14.)	0,003 (16.)	0,001 (18.)	0,001 (20.)
0,7	0,17 (11.)	0,10 (13.)	0,05 (15.)	0,03 (17.)	0,02 (20.)	0,01 (22.)	0,004 (25.)	0,002 (27.)
0,8	0,24 (16.)	0,15 (18.)	0,10 (20.)	0,06 (23.)	0,03 (26.)	0,02 (29.)	0,01 (32.)	0,01 (35.)
0,9	0,35 (20.)	0,25 (22.)	0,17 (25.)	0,11 (29.)	0,08 (32.)	0,05 (36.)	0,03 (40.)	0,02 (44.)
1,0	0,52 (24.)	0,40 (27.)	0,30 (31.)	0,22 (35.)	0,17 (39.)	0,13 (43.)	0,09 (48.)	0,07 (50.)

$$\varepsilon(\alpha, \tau, B, \lambda_{\max}) = U(\alpha, B, \lambda_{\max}) + V(\tau, B, \lambda_{\max}),$$

где $B = \langle B^t \rangle + B^a$; $\langle B^t \rangle$ – среднее значение температурного фактора, а оценки функций $U(\alpha, B, \lambda_{\max})$ и $V(\tau, B, \lambda_{\max})$ даны в (13) и (18).

Как уже упоминалось, изменение параметра B^a практически не влияет на время расчета. Поэтому при заданных α и τ целесообразно выбирать B^a так, чтобы при этом значении B^a ошибка $\varepsilon(\alpha, \tau, B, \lambda_{\max})$ была минимальна. Мы определим оптимальное значение $B_{\text{opt}} = B_{\text{opt}}(\alpha, \tau, \lambda_{\max})$ как значение B , при котором функция $\varepsilon(\alpha, \tau, B, \lambda_{\max})$ достигает минимума:

$$B_{\text{opt}}(\alpha, \tau, \lambda_{\max}) = \arg \min_B \varepsilon(\alpha, \tau, B, \lambda_{\max}),$$

и определим наилучшую точность, достигаемую при данных α и τ , как

$$\varepsilon_{\text{opt}}(\alpha, \tau, \lambda_{\max}) = \varepsilon(\alpha, \tau, B_{\text{opt}}, \lambda_{\max}) = \min_B \varepsilon(\alpha, \tau, B, \lambda_{\max}).$$

Значение B_{opt} определяет соответствующее значение добавочного температурного фактора $B_{\text{opt}}^a = B_{\text{opt}} - \langle B^t \rangle$.

В качестве примера в табл. 3 приведены значения наилучшей достижимой точности ε_{opt} и оптимальное значение B_{opt} при различных значениях α и τ в случае $\lambda_{\max} = 0,5$ (то есть когда рассчитывается набор структурных факторов, соответствующий разрешению 2 Å).

2. Из табл. 3 видно, что заданная точность расчета набора структурных факторов может быть достигнута при различных комбинациях параметров α и τ . Естественно пытаться при заданной точности расчета выбрать параметры α и τ так, чтобы сделать время расчета минимальным. Если нам известна зависимость времени расчета набора структурных факторов от величин α и τ – функция $T(\alpha, \tau)$, то мы можем определить оптимальный радиус атома α_{opt} и оптимальный шаг сетки τ_{opt} как значения α и τ , при которых достигается минимальное значение $T(\alpha, \tau)$ при условии, что $\varepsilon_{\text{opt}}(\alpha, \tau) \leq \varepsilon_0$, где ε_0 – заданная точность.

Заключение

В заключение отметим, что все проведенные выше рассмотрения носили приближенный характер. Поэтому полученные оценки точности не носят абсолютного характера, а лишь определяют область разумных значений параметров. Действительную точность по-

лучаемых при этих значениях параметров структурных факторов можно проверить, сравнив их (для не слишком большого случайного набора рефлексов), со структурными факторами, вычисленными по точной формуле (1).

ЛИТЕРАТУРА

1. Бландел Т., Джонсон Л. Кристаллография белка. М., Мир, 1979.
2. Сринивасан Р., Паргасарати С. Применение статистических методов в рентгеновской кристаллографии. М., Мир, 1979.
3. Ten Eyck L.F. - Acta Cryst., 1977, v. A33, p. 486-492.
4. Ten Eyck L.F. - Acta Cryst., 1973, v. A29, p. 183-191.
5. Agarwal R.C. - Acta Cryst., 1978, v. A34, p. 791-809.
6. Schmid M.F., Herriott J.R. - J. Mol. Biol., 1976, v. 103, p. 175-190.

СОДЕРЖАНИЕ

Введение	1
§ 1. Алгоритм расчета структурных факторов	2
§ 2. Влияние "радиуса" атома на точность расчета структурных факторов	8
§ 3. Влияние величины шага сетки на точность расчета структурных факторов	10
§ 4. Оптимальный выбор параметров метода	13
Заключение	15

Владимир Юрьевич Лунин

ОПТИМИЗАЦИЯ РАСЧЕТА СТРУКТУРНЫХ ФАКТОРОВ
В КРИСТАЛЛОГРАФИИ БЕЛКА. Препринт

Отредактировано и подготовлено к печати в ОНТИ НЦБИ

Редактор Г.П.Медведева

Технический редактор С.М.Ткачук

Корректоры В.И.Дубровина, Л.М.Орлова

Подписано в печать 14.06.82 г.

Т09479. Уч.-изд.л. 1,1. Формат 60x60/16.

Тираж 100 экз. Бумага офсетная. Заказ 2379Р.

Бесплатно. Изд. № 205.

Отпечатано на ротапринте в Отделе научно-технической информации
Научного центра биологических исследований АН СССР
в Пушкине